**TS 18661 DR 19**

=============================================

**Submitter:** C FP group
**Submission Date:** 2018-03-16
**Source:** WG14
**Reference Documents:** TS 18661-1, C11
**Subject:** updating underflow definition

**Summary**

C11 7.12.1 says

> The result underflows if the magnitude of the mathematical result is so small that the mathematical result cannot be represented, without extraordinary roundoff error, in an object of the specified type.232)

> 232) The term underflow here is intended to encompass both "gradual underflow" as in IEC 60559 and also "flush-to-zero" underflow

The C definition of underflow doesn't accomplish its intention as expressed in footnote 232. It's too restrictive to encompass IEC 60559 gradual underflow, because IEC 60559 underflow can occur without extraordinary roundoff error.

In IEC 60559:2011, the underflow flag is raised (in C terminology, set) if and only if the result is tiny and inexact, even if the roundoff error is the same as it would have been if the full (normal) precision of the type were available, i.e., even if there is no extraordinary roundoff error.

IEC 60559:1989 offered the implementation the option to raise the underflow flag for tiny results with extraordinary roundoff error, though it also offered the option to raise the underflow flag for tiny inexact results, which is what most implementations did. IEC 60559:2011 dropped the option to raise the underflow flag based on extraordinary roundoff error.

The following suggested TC aims to loosens the C definition of underflow to encompass IEC 60559:2011 underflow behavior, which is based on tiny inexact results.

**Suggested Technical Corrigendum**

In TS 18661-1, before 14.1, insert:

## 14.0 C underflow

The following change to C11 loosens the C definition of underflow to encompass IEC 60559 gradual underflow, as is its stated intention (see C11 footnote 232).

**Changes to C11:**

Change the first sentence in 7.12.1#6 from:

[6] The result underflows if the magnitude of the mathematical result is so small that the mathematical result cannot be represented, without extraordinary roundoff error, in an object of the specified type.232) …

to:

[6] The result underflows if the magnitude of the mathematical result is nonzero and less than the minimum normal number in the type.232) …